Extended proof of Annex B.2 of Neural Ordinary differential equation

Samuel Prevost <samuel.prevost@etu.upmc.fr>

August 17, 2019

Let

1. $f(z(t), \theta(t), t) = \frac{dz(t)}{dt}$
2. $Z = \begin{bmatrix} z(t), & \theta(t), & t \end{bmatrix}$
3. $a(t) = \frac{dL}{dZ(t)} = \begin{bmatrix} \frac{dL}{dz(t)}, & \frac{dL}{d\theta(t)}, & \frac{dL}{dt} \end{bmatrix} = \begin{bmatrix} a_z(t), & a_\theta(t), & a_t(t) \end{bmatrix}$
4. $g(Z(t), t) = \frac{dZ(t)}{dt} = \begin{bmatrix} \frac{dz(t)}{dt}, & \frac{d\theta(t)}{dt}, & \frac{dt}{dt} \end{bmatrix} = \begin{bmatrix} f, & 0_{\mathbf{E}}, & 1 \end{bmatrix}$ where $0_{\mathbf{E}}$ is a null matrix of appropriate size

Then, we can define

$$Z(t+\varepsilon) = Z(t) + \int_t^{t+\varepsilon} g(Z(t), t)dt = T_\varepsilon(Z(t), t)$$

notice that

$$a(t) = \frac{dL}{dZ(t)} = \frac{dL}{dZ(t+\varepsilon)} \frac{dZ(t+\varepsilon)}{dZ(t)} = a(t+\varepsilon) \frac{\partial T_\varepsilon(Z(t), t)}{\partial Z(t)}$$

Thus we can derive $a(t)$:

$$
\begin{aligned}
\frac{da(t)}{dt} &= \lim_{\varepsilon \to 0^+} \frac{a(t+\varepsilon) - a(t)}{\varepsilon} \\
&= \lim_{\varepsilon \to 0^+} \frac{a(t+\varepsilon) - a(t+\varepsilon)\frac{\partial}{\partial Z(t)} T_\varepsilon(Z(t), t)}{\varepsilon} \\
&= \lim_{\varepsilon \to 0^+} \frac{a(t+\varepsilon) - a(t+\varepsilon)\frac{\partial}{\partial Z(t)} \left(Z(t) + \varepsilon g(Z(t), t) + \mathcal{O}\left(\varepsilon^2\right)\right)}{\varepsilon} \\
&= \lim_{\varepsilon \to 0^+} \frac{a(t+\varepsilon) - a(t+\varepsilon) \left(I + \varepsilon \frac{\partial g(Z(t),t)}{\partial Z(t)} + \mathcal{O}\left(\varepsilon^2\right)\right)}{\varepsilon} \\
&= \lim_{\varepsilon \to 0^+} \frac{-\varepsilon a(t+\varepsilon)\frac{\partial g(Z(t),t)}{\partial Z(t)} + \mathcal{O}(\varepsilon^2)}{\varepsilon} \\
&= \lim_{\varepsilon \to 0^+} -a(t+\varepsilon)\frac{\partial g(Z(t),t)}{\partial Z(t)} + \mathcal{O}(\varepsilon) \\
&= -a(t)\frac{\partial g(Z(t),t)}{\partial Z(t)}
\end{aligned}
$$

As you can see, the proof of appendix B.1 holds for any vector Z. What's more interesting is the way we can exploit this to get $\frac{da_\theta(t)}{dt}$:

$$
-a(t)\frac{\partial g(Z(t),t)}{\partial Z(t)} = \begin{bmatrix} a_z(t), & a_\theta(t), & a_t(t) \end{bmatrix} \cdot \begin{pmatrix} \frac{df(z(t),\theta(t),t)}{dz} & \frac{df(z(t),\theta(t),t)}{d\theta} & \frac{df(z(t),\theta(t),t)}{dt} \\ \frac{d\theta/dt}{dz} & \frac{d\theta/dt}{d\theta} & \frac{d\theta/dt}{dt} \\ \frac{dt/dt}{dz} & \frac{dt/dt}{d\theta} & \frac{dt/dt}{dt} \end{pmatrix}
$$

But as mentioned in the paper, $\frac{d\theta(t)}{dt} = 0$ and $\frac{dt}{dt} = 1$ which means that the second and third rows are null (derivatives of constants). Hence:

$$
-a(t)\frac{\partial g(Z(t),t)}{\partial Z(t)} = \begin{bmatrix} a_z(t), & a_\theta(t), & a_t(t) \end{bmatrix} \cdot \begin{pmatrix} \frac{df}{dz} & \frac{df}{d\theta} & \frac{df}{dt} \\ 0_{\mathbf{E}} & 0_{\mathbf{E}} & 0_{\mathbf{E}} \\ 0_{\mathbf{E}} & 0_{\mathbf{E}} & 0_{\mathbf{E}} \end{pmatrix} =
$$

$$
- \begin{bmatrix} a_z(t)\frac{df}{dz} + a_\theta 0_{\mathbf{E}} + a_t 0_{\mathbf{E}} \\ a_z(t)\frac{df}{d\theta} + a_\theta 0_{\mathbf{E}} + a_t 0_{\mathbf{E}} \\ a_z(t)\frac{df}{dt} + a_\theta 0_{\mathbf{E}} + a_t 0_{\mathbf{E}} \end{bmatrix} = - \begin{bmatrix} a_z(t)\frac{df}{dz} \\ a_z(t)\frac{df}{d\theta} \\ a_z(t)\frac{df}{dt} \end{bmatrix}
$$

Now, we've got:

$$\frac{da(t)}{dt} = -\left[a_z(t)\frac{df}{dz}, \quad a_z(t)\frac{df}{d\theta}, \quad a_z(t)\frac{df}{dt}\right] = -\frac{d}{dt}\left[\frac{dL}{dz(t)}, \quad \frac{dL}{d\theta(t)}, \quad \frac{dL}{dt}\right]$$

Hence:

$$\frac{d}{dt}\left(\frac{dL}{d\theta(t)}\right) = -a_z(t)\frac{df}{d\theta}$$

$$\implies \frac{dL}{d\theta(t)} = -\int a_z(t)\frac{df}{d\theta}dt$$